RF ROTHWELL FIGG

901 New York Avenue, N.W.
Suite 900 East
Washington, DC 20001

Telephone (202) 783-6040
Facsimile (202) 783-6031
www.rothwellfigg.com
info@rothwellfigg.com

Offices in Washington, D.C. | New York | Boston

Hon. Ona T. Wang
United States Magistrate Judge
Southern District of New York

<div align="right">

G. Franklin Rothwell
(1928-2011)

E. Anthony Figg	Jennifer B. Maisel
George R. Repper	Jess M. Collen
Steven Lieberman	Jeffrey A. Lindenbaum*
Joseph A. Hynds	Mark T. Rawls
Martin M. Zoltick	Monica Chin Kitts
R. Danny Huntington	Nicole M. DeAbrantes
Joo Mee Kim	Davide F. Schiavetti
Brian S. Rosenbloom	Vivian Y. Tian
Jennifer P. Nock	Kristen J. Logan
Brett A. Postal	Sheena X. Wang
Aydin H. Harston, Ph.D.	Gary J. Prato
Richard Wydeven	An Nguyen
Sharon L. Davis	James T. Pawlowski
Robert P. Parker	Bryan B. Thompson
Jenny L. Colgate	Melissa C. Santos, Ph.D.
Leo M. Loughlin	Andrew C. Stewart
Michael V. Battaglia	Adam M. Nicolais
Michael H. Jones	Gregory S. DeLassus
Rachel M. Echols	Michael P. Ellenberger*
Daniel L. Shores	

*Not a member of the D.C.
Bar. Practice limited to patent,
trademark and copyright matters
and cases in federal courts.

Of Counsel
Harry F. Manbeck, Jr.
Jane F. Collen
James R. Hastings*
Mitchell E. Radin*

</div>

Re: *The New York Times Company v. Microsoft Corporation, et al.*, 23-cv-11195; *Daily News, L P, et al. v. Microsoft Corp., et al.*, 1:24-cv-3285

Dear Magistrate Judge Wang:

I write on behalf of Plaintiffs Daily News, LP et al. ("Daily News") and The New York Times Company[1] ("The Times") (collectively, the "News Plaintiffs"). We oppose OpenAI's request[2] for a protective order prohibiting disclosure of OpenAI Confidential Information to Dr. Baeza-Yates. *See* Dkt. 163.

1. There is no risk of commercial harm to OpenAI because Theodora.AI is not OpenAI's competitor

OpenAI's motion never states that Theodora.AI is an actual or potential competitor to OpenAI. When asked whether Theodora.AI is a competitor to OpenAI in the parties' meet and confer, OpenAI would not answer. Rawls Decl. ¶ 2. The simple answer is—it is not. Baeza-Yates Decl. ¶ 3. Theodora.AI is a small, six-person Chilean company that provides recommendations about how to fix bias issues (such as gender bias) in Spanish-language documents. *Id.* ¶ 3. It does not pre-train LLMs, does not use retrieval-augmented generation, and instead only fine-tunes the LLMs of others for the purpose of detecting and fixing such bias issues. *Id*. ¶¶ 4-6. OpenAI's motion is silent as to the extent to which OpenAI focuses its fine-tuning efforts on mitigating Spanish-language bias, if any, or why this case would necessitate disclosure of any documents related to it. Indeed, in the context of GPT-4 post-training safety mitigations, OpenAI has publicly stated that "[m]itigations and measurements were mostly designed, built, and tested primarily in English and with a US-centric point of view. The majority of pretraining data and our alignment data is in English."[3]

OpenAI's motion fails because there is no risk of competitive harm. Absent Theodora.AI being a competitor (which it is not), OpenAI cannot establish any harm to its "interest in protecting

---

[1] While OpenAI's objection is directed to the *Daily News* Plaintiffs, the New York Times also recently disclosed Dr. Baeza-Yates as an expert under the protective order, and joins in this response.

[2] Notably, OpenAI brings this motion alone. Its co-defendant, Microsoft, does not object to Dr. Baeza-Yates receiving its highly confidential materials, presumably because it recognizes that Theodora.AI is not a competitor.

[3] *See* OpenAI, GPT-4 SYSTEM CARD (Mar. 10, 2023), https://cdn.openai.com/papers/gpt-4-system-card.pdf.

RF ROTHWELL FIGG

Hon. Ona T. Wang
October 22, 2024
Page 2

[its highly confidential] materials from a potential *competitor*." *Flores v. Stanford*, No. 18-cv-02468, 2021 WL 4441614, at *4 (S.D.N.Y. Sep. 28, 2021) (emphasis added); *see also Infosint S.A. v. H. Lundbeck A.S.*, No. 06-cv-2869, 2007 WL 1467784, at *2 (S.D.N.Y. May 16, 2007) (noting that "the [producing] party's ability to protect its confidential materials from misuse by *competitors*" is "evaluated" against the other "party's right to broad discovery") (emphasis added, citation and quotation marks omitted).

At most, OpenAI argues that the companies "operate[] within the same space." Dkt. 163 at 2. But this space— "the area of developing AI-based tools," as OpenAI defines it—is too broad to be workable. *Id.* General Motors and a car-wash company both operate in the automotive space, but that does not make them competitors. Moreover, simply being "a current consultant in the industry with experience and knowledge about the subject matter at issue …is, to a certain extent, a prerequisite for participation in this litigation and not grounds for disqualification." *See First Quality Tissue, LLC v. Irving Consumer Prods. Ltd.*, No. 19-cv-428, slip op. at 21 (D. Del. Oct. 9, 2020); *see also Advanced Micro Devices, Inc. v. LG Elecs., Inc.*, No. 14-cv-01012, 2017 WL 3021018, at *3 (N.D. Cal. July 17, 2017) ("The Court finds it unlikely that AMD would be able to select an expert in this space who does not consult, publish, teach, or invent in the relevant field.").

2. The relevant balancing factors favor the News Plaintiffs

As a threshold matter, the *Flores* balancing test is inapplicable to this dispute because Theodora.AI is not a competitor to OpenAI. *See Flores*, 2021 WL 4441614, at *4 (stating the test weighs a "party's need to have its chosen expert review the materials against the producing party's interest in protecting the materials from a potential *competitor*." (emphasis added)). But even if it applied, each factor favors the News Plaintiffs, not OpenAI.

First, Dr. Baeza-Yates is not "involved in *competitive* decision-making"—because Theodora.AI is not a competitor of OpenAI. This precludes a finding of any commercial harm to OpenAI from Dr. Baeza-Yates receiving OpenAI's highly confidential information under the terms of the protective order. *See, e.g.*, *Advanced Micro Devices*, 2017 WL 3021018, at *3 (allowing experts to receive highly confidential material where neither expert worked for or assisted in competitive decision-making at a competitor of the opposing party).

Second, there is no risk of inadvertent disclosure of proprietary information. OpenAI acknowledges that this risk "is typically triggered 'when the proposed recipient of sensitive information has an active relationship with one or more of the disclosing party's *competitors*, and is involved in *competitive* decision-making." Dkt. 163 at 3 (emphasis added), *quoting Flores*, 2021 WL 4441614, at *9. Here, there is no relationship with a competitor. OpenAI also fails to specifically identify what the risk of inadvertent disclosure is, or to provide specificity about what kinds of material Dr. Baeza-Yates would be unable to keep confidential.

OpenAI goes further—alleging not just an inadvertent, accidental disclosure of highly confidential information, but actually suggesting to the Court that "there is a *guarantee* of *advertent* [i.e., intentional] disclosure directly to Theodora AI via Dr. Baeza-Yates." Dkt. 163 at 3. OpenAI provides no basis for supporting this statement, and there is none. Dr. Baeza-Yates has signed the

RF ROTHWELL FIGG

Hon. Ona T. Wang
October 22, 2024
Page 3

protective order, and there is no reason to doubt his ability to keep the information he learns confidential. OpenAI has no basis to impugn Dr. Baeza-Yates's honesty and professionalism. This Court "should not lightly assume that" Dr. Baeza-Yates "will violate [the] obligations imposed on [him] by a duly entered protective order on the court's docket." *Flores*, 2021 WL 4441614 at *8 (cleaned up); *Santella v. Grizzly Indus., Inc.*, No. 12-mc-00131, 2012 WL 5399970, at *5 (D. Or. Nov. 5, 2012) ("[T]he Court concludes that Mr. Domeny will abide by the protective order, which negates the risk of Mr. Domeny intentionally disclosing SawStop's confidential information."); *Nellson N. Operating, Inc. v. Elan Nutrition, LLC*, 238 F.R.D. 544, 547 (D. Vt. 2006) (declining to doubt the expert's undertaking to keep the opposing party's information confidential).

Third, the hardship imposed by the proposed restriction is that it would deprive the News Plaintiffs of the benefit of having their chosen expert review material directly relevant to the asserted claims and would impede Dr. Baeza-Yates's ability to give effective testimony if he is ultimately designated as a testifying expert. Given the technological complexities of this case, experts are essential. *See Nellson N. Operating, Inc.*, 238 F.R.D. at 547 (holding that "[e]xpert testimony will be essential to understanding the parties' claims in this case and Elan must have latitude to choose the experts that it believes have the necessary expertise and ability to convey highly technical information to the Court."). While OpenAI refers to the presence of other technical experts in the case, Dr. Baeza-Yates has an extraordinary background in information retrieval (for example, through his work at Yahoo) that makes him particularly well-suited to address issues related to retrieval-augmented generation, among other issues in the case. Although the other experts work broadly in "computer-related fields," it is not true that Dr. Baeza-Yates's expertise "overlap[s] entirely" with the expertise of the other technical experts.[4] Dkt. 163 at 3. *See Nellson N. Operating, Inc.*, 238 F.R.D. at 547 (recognizing that the presence of another expert did not mitigate the harm of denying access of confidential information to an additional expert).

Fourth, as to the timing of the remedy, discovery is well underway; the News Plaintiffs have begun inspecting OpenAI's training datasets and expect to begin inspecting source code and other technical material soon. Given the lack of any competitive harm, there is no reason to bar the News Plaintiffs' preferred expert from receiving relevant, highly confidential information precisely when his input could aid the News Plaintiffs in efficiently litigating their case.

Fifth, OpenAI's proposed remedy (even if it is justified, which it is not) is plainly overbroad. Theodora.AI does not develop foundational models, it does not pre-train LLMs, and it does not use retrieval-augmented generation techniques. Discovery into all these areas is highly relevant to the asserted claims. Yet, OpenAI's proposed remedy would prevent Dr. Baeza-Yates from accessing highly confidential material on any of these issues, even though there is no basis to suggest any risk of inadvertent disclosure of them to Theodora.AI.

For these reasons, the Court should refuse OpenAI's request for a protective order prohibiting disclosure of OpenAI Confidential Information to Dr. Baeza-Yates.

---

[4] Vinay Hooloomann, identified by OpenAI as an alternative to Dr. Baeza-Yates, is a source code reviewer. Presumably so are the five other unidentified individuals mentioned in Ms. Nightingale Dawson's Declaration.

ROTHWELL FIGG

Hon. Ona T. Wang
October 22, 2024
Page 4

October 22, 2024   Respectfully submitted,

                                                      /s/Steven Lieberman
                                                      Steven Lieberman

                                                      /s/Ian B. Crosby
                                                      Ian B. Crosby

cc: All Counsel of Record (via ECF)